CSC/DSP 310: Programming for Data Science

Syllabus – Spring 2026

Time:

CSC/DSP 310 Section 1 MWF 10-10:50, Location: Bliss Hall room 290
CSC/DSP 310 Section 2 MWF 11-11:50, Location: Washburn Hall 208

Webpage: BrightSpace and https://lutzhamel.github.io/CSC310
Prerequisites: CSC201 or CSC211

Instructor:
Prof. Lutz Hamel
email: lutzhamel@uri.edu
office: Tyler Hall Rm 251
office hours: (see BrightSpace)

Course Description

Data science exists at the intersection of computer science, statistics, and machine learning. That means writing programs to access and manipulate data so that it becomes available for analysis using statistical and machine learning techniques is at the core of data science. Data scientists use their data and analytical ability to find and interpret rich data sources; manage large amounts of data despite hardware, software, and bandwidth constraints; merge data sources; ensure consistency of datasets; create visualizations to aid in understanding data; build mathematical models using the data; and present and communicate the data insights/findings.

This course provides a survey of data science. Topics include data driven programming in Python; data sets, file formats and meta-data; descriptive statistics, data visualization, and foundations of predictive data modeling and machine learning; accessing web data and databases; distributed data management. You will work on weekly substantial programming problems such as accessing data in database and visualize it or build machine learning models of a given data set.

Goals

The primary aim of CSC 310 is to introduce you to programming in the context of data science and statistical thinking by providing a survey of the major technologies and techniques that are currently being employed.
The objectives of CSC 310 are

- To provide an introduction to data sets, file formats, and meta-data.
- To provide an introduction to database systems such as MySQL.
- To provide a basic overview of data manipulation, statistical data summary techniques, and visualization.
- To provide an introduction to data modeling techniques, in particular computational techniques usually referred as “machine learning”.

Learning Outcomes

At the end of this course, students will be able to:

- Describe what data science is with a detailed view
- Access and visualize data
- Build and evaluate models of data
- Solve problems in data science using standard tools and techniques

Texts

Python Data Science Handbook, Jake VanderPlas, O'Reilly, 2017. (available free online)

Software

We will be using Google's Colab Notebooks available with every Google account including URI student accounts.

Laptops

You will need to bring a working, fully charged laptop to class for in-class lab projects.

Grading

In-Class Lab Projects	20%
Midterm Project	40%
Final Project	40%

IMPORTANT: To receive full credit for **lab/midterm/final projects** you will have to be **physically present** in the classroom at the time of these projects. **A 20% penalty will be applied to work submitted without having been physically present during these projects**.

Grading Key

Symbol*	Start %*
F	0
D	60
D+	67
C-	70
C	73
C+	77
B-	80
B	83
B+	87
A-	90
A	93

Policies

- Check BrightSpace!
- **Email Policy:** Emails sent to me on **Monday through Friday between 9am and 5pm** will usually be answered within an hour or two. For emails sent to me outside of these times you can expect an answer on the next business day.
- Class **promptness, participation,** and **adequate preparation** for each class are expected. If you are absent, it is your responsibility to find out what you missed (e.g. handouts, announcements, assignments, new material, etc.)
- **Late submissions** will not be accepted.
- All work is to be the result of your own individual efforts unless explicitly stated otherwise. **Plagiarism, unauthorized cooperation, or any other form of cheating including the copying code from ChatGPT (or similar) will be reported to the Dean**. See the appropriate sections (8.27) of the University Manual.
- Any student with a documented disability is welcome to contact me as early in the semester as possible so that we may arrange accommodations. As part of this process, please be in touch with Disability Services for Students Office at 302 Memorial Union, Phone 401-874-2098.

Tentative Schedule

Week 1 & 2

- What is Data Science?
- Data Literacy

- Jupyter Notebooks

Week 3

- A Quick Crash Course to get You started
    - Data Sources
        - The CSV File
    - Basic Descriptive Statistics
    - Visualizing Data
    - Model Building/Machine Learning

Week 4 & 5

- More on working with Numeric Data
    - Pandas Data Frames
    - Data Cleaning/Transforming
- Data Visualization with Pandas and other Python Modules

Week 6 & 7

- Models of Data
    - Statistical Models
    - Machine Learning
- Evaluating Models
    - Model Selection
    - Model Comparisons
- Model Deployment

Week 8

- Natural Language/Text Processing (NLP)
    - Bag of Words
    - n-Grams

Week 9

- Databases
    - MySQL

Week 10

- Deep Learning
    - Convolutional networks for image processing
    - Deep neural networks for NLP

Tentative In-Class Lab Schedule

1. Jupyter Notebooks
2. Data Set Basics - the CSV File: loading, manipulating and summary statistics
3. Machine Learning: Decision Trees
4. Model Evaluation I

5. Working with Python Data Frames
6. Data Visualization
7. Model Evaluation II: Cross-Validation, Confusion Matrices, and Confidence Intervals
8. Additional ML Models: KNN and ANN
9. NLP: “Fake News Detection”